DOCUMENT RESUME

ED 316 932 EA 021 712

AUTHOR Baron, Joan Boykoff; And Others
TITLE Toward a New Generation of Student Outcome Measures: Connecticut's Common Core of Learning Assessment.
PUB DATE Mar 89
NOTE 44p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Educational Assessment; Educational Improvement; Elementary Secondary Education; *Evaluation Criteria; *Evaluation Methods; *Evaluation Utilization; *Formative Evaluation; Outcomes of Education; *State Programs
IDENTIFIERS Common Core of Learning; *Connecticut State Department of Education

ABSTRACT

A description of ecologically valid, authentic, and integrated performance-based assessment instruments of different lengths was undertaken to measure the extent by which students meet Connecticut's Common Core of Learning (CCL) objectives. The description focuses on the students' knowledge and general understanding of science and mathematics; general communicative, quantitative, and critical thinking skills; and relevant attitudes and dispositions. Seven tasks will be accomplished to implement the assessment: (1) define essential understandings in science and mathematics; (2) select a statewide sample and solicit participation of local option schools; (3) plan and conduct the Summer Institutes for Teaching and Learning (ITL) workshops; (4) identify psychometric issues and considerations related to the validity, reliability, and practicality of the assessment instruments; (5) implement performance tasks and scoring rubrics; (6) report on the condition of science and mathematics education; and (7) release performance tasks and scoring rubrics for use in Connecticut classrooms and elsewhere. Once the assessment is implemented, it is hoped that students will construct meaning from the results. (JAM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED316932

Toward a New Generation of Student Outcome Measures: Connecticut's Common Core of Learning Assessment

by

Joan Boykoff Baron
Pascal D. Forgione, Jr.
Douglas A. Rindone
Hannah Kruglanski
and
Bruce Davey

Connecticut State Department of Education
Division of Research, Evaluation and Assessment
165 Capitol Avenue - Room 340
Hartford, Connecticut 06106
(203) 566-4001

Presented at

American Educational Research Association
Annual Meeting
San Francisco, California

Part of

Symposium 28.01
Beyond Effective Schools:
Quality Indicators for Evaluating Schools and Districts

March 29, 1989

EA 021 712

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Douglas Rindone

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TABLE OF CONTENTS

A. Overview of Connecticut's Common Core of Learning (CCL) Assessment Development Project 1

B. The Need for Alternative Assessment Approaches in Science and Mathematics 2
- The Condition of Science Education and Assessment
- The Condition of Mathematics Education and Assessment
- Connecticut's Strategy for Meeting the Critical Needs

C. Connecticut's Capacity and Commitment to Developing Performance Testing in Science and Mathematics 6
- Connecticut's Prior Experiences in Performance Testing
- Developing New Indicators of Science and Mathematics Achievement Through Performance Testing
- Connecticut's Commitment to Innovative Performance Assessments
- Preliminary Blueprint for Prototype Performance Assessments
- Collaboration with the Coalition for Essential Schools

D. Major Common Core of Learning Project Tasks and Activities . . 13
- Task 1: Defining the Essential Understandings in Science and Mathematics
- Task 2: Selecting a Statewide Sample and Soliciting Participation of Local School Options
- Task 3: Planning and Conducting the Summer Institutes for Teaching and Learning (ITL) Workshops
- Task 4: Identifying Psychometric Issues and Considerations Related to the Validity, Reliability, and Practicality of these Instruments
- Task 5: Implementing the Performance Tasks and Rubrics
- Task 6: Reporting on the Condition of Science and Mathematics Education, and
- Task 7: Releasing Performance Tasks, Rubrics and Handbooks for Use in Classrooms in Connecticut and in the Nation

E. Guidelines for Development and Use of Statewide Common Core of Learning Assessments in Grades 9-12 19
- I. Rationale for Common Core of Learning (CCL) Assessment
  - Common Core of Learning Assessment Overview
- II. Specifications for Performance Tasks
  - Prompt Statements
  - Desirable Task Characteristics
  - Task Categorization
- III. Specifications for Scoring Rubrics
  - The Scoring Paradigm
  - Rating System Properties
  - Identifying Performance Dimensions to be Scored
  - Provision for Student Feedback
- IV. Early Development Work

Bibliography of Pertinent Literature 28

Attachment 1: Sample Mathematics Prompt 31

A. Overview of Connecticut's Common Core of Learning (CCL) Assessment Development Project

The Connecticut State Department of Education (CSDE) has begun a multi-year assessment of the extent to which students in Connecticut display the knowledge and traits described in the Common Core of Learning (CCL, see Connecticut State Board of Education, 1987). The CCL is a document which includes a comprehensive set of learning outcomes, including many which require higher order thinking, interpersonal competency and a variety of personal traits. In sum, the CCL portrays a statement of Connecticut's high expectations for a broad range of student outcomes resulting from the K-12 school experience.

Science and mathematics were selected as the first two subjects to be assessed under the CCL Assessment Project because of their critical importance to Connecticut's highly technological and technically oriented economy and because of the CSDE's in-depth experience and expertise in assessing these two areas. These areas are also receiving simultaneous attention in Connecticut's Teacher Assessment Project (see Pecheone, Baron, Forgione and Abeles, 1988). The CSDE is currently in a developmental year of a comprehensive assessment of these two areas in grades 9 through 12. This is a highly innovative project which is designed to provide a much needed alternative to multiple-choice testing. Connecticut will design and implement a series of ecologically valid, authentic, and integrated performance-based assessment instruments of different lengths to assess students' knowledge and understandings of science and mathematics, as well as students' general communicative, quantitative and critical thinking skills, and relevant attitudes and dispositions. This developmental effort has important implications for the integration of curriculum, instruction, and evaluation.

By providing successful models of assessment that involve sustained thinking and problem-solving, the participants in this collaborative effort hope to stimulate a new genre of teaching and classroom assessment devices that includes exhibitions, hands-on performance experiences, student portfolios, and other divergent tasks that require students to actively engage in sustained exploration. The CCL assessment tasks will require students to work individually and in groups to solve complex, multi-step problems, and design and carry out investigations that require students to frame problems, collect data, and analyze, integrate, interpret, and report their results to real audiences. Imbedded in the tasks will be opportunities for students to construct meaning and structure investigations. Furthermore, they will be asked to reflect upon their experiences, in order to reach deeper levels of understanding about the content and practice of science and mathematics and communicate those new understandings to others.

In developing and pilot testing performance tasks and scoring rubrics, the CSDE is looking forward to working collaboratively with teachers in six other states (i.e., Michigan, Minnesota, New York, Texas, Vermont and Wisconsin), the leadership and teachers in the Coalition of Essential Schools, and a team of national consultants in education, mathematics, science and psychology. This will ensure that our efforts to develop performance assessments are generalizable to a national audience.

B. The Need for Alternative Assessment Approaches in Science and Mathematics

In both science and mathematics, recent reports and assessments have made clear the pressing need for alternative approaches for both teaching and assessing what students know and can do in these two critical subject areas. The Nation's Report Card, released on February 14, 1989 (Roberts, 1989), lamented the fact that "too few students develop the capacity to use the knowledge and skills they require in school for thoughtful or innovative purposes." The report went on to elaborate that "too few students learn to reason effectively about information from the subjects they study." It is clear that in order to improve students' knowledge and understanding of science and mathematics, tomorrow's instruction and assessments must be different from today's in profound ways (see IEA, 1988; McKnight, Crosswhite, Dossey, Kiffer, Swafford, Travers and Cooney (1987); and Mullis and Jenkins, 1988).

The Condition of Science Education and Assessment

The June 22, 1988 comments by Harmon, Mokros, Dawson, Hartwig, Henderson, Lowery, and Taylor on the need for new assessment materials was confirmed at two recent meetings sponsored by the National Science Foundation, in November 1988, in Chatham, Massachusetts (see Hein, 1988) and in January, 1989 in Tampa, Florida (see Blank, 1989). Science educators and measurement experts at both meetings strongly emphasized the national need for new and expanded examples of performance testing in our nation's schools. A report soon to be released by the National Center for the Improvement of Science Education on assessing science in elementary school will reiterate this point (see Raizen, Baron, Champagne, Haertel, Mullis, and Oakes, 1989). A key point made by Harmon, et.al. is that an over-reliance on multiple-choice testing not only emphasizes simple recall of facts and/or recognition of textbook experiments, but militates against the less predictable hands-on approach. They also state that "the existing norm-referenced tests not only fail to support or encourage the implementation of new developments in science curriculum and pedagogy, but their continued, near-universal use may dampen or totally inhibit implementation of such approaches." This claim finds support in data from the NAEP Report Card showing that the percentage of students in high school (age 17) who were at or above proficiency levels on higher order thinking skills in science were shockingly low. Only 7% can infer relationships and draw conclusions using detailed knowledge and 41% have some detailed scientific knowledge and can evaluate the appropriateness of scientific procedures.

The Condition of Mathematics Education and Assessment

A parallel set of reports is currently being released in mathematics. The current year, 1989, is likely to be known as the "Year of the Mathematics Reform Reports." To some extent, these reports may be seen as a direct response to the very alarming finding reported by NAEP "that only 6% of our seventeen-year-olds can solve multi-step problems and use basic algebra (see Dossey, Mullis, Lindquist, and Chambers, 1988). This year has already seen the release of Everybody Counts developed by the Mathematics and Sciences Education Board (MSEB) of the National Academy of Sciences (see National Research Council, 1989). Soon to be released are the Curriculum and

Evaluation Standards for School Mathematics from the National Council of Teachers of Mathematics (see NCTM Working Draft, 1987) and the Mathematics Framework K-12 (MSEB). The overarching message of these documents is parallel to that in the Project 2061 Report released by the American Association for the Advancement of Sciences (1988), i.e., that what is needed is "a shift in emphasis from a curriculum dominated by an emphasis on memorization of isolated facts and procedures, and proficiency with paper-and-pencil skills to one which emphasizes conceptual understandings, multiple representations and connections, mathematical modeling, and mathematical problem solving." Mathematicians are stressing that "learning is not a matter of collecting, but of constructing" with the expectation that students will construct, investigate, translate, discuss, analyze and formulate mathematics in a dynamic way. In the NCTM Standards section entitled "Topics to Receive Increased Attention" are recommendations for using: real-world problems to motivate students to apply theory; computer graphing to develop conceptual understanding; integration across and connections among topics at all grade levels; and realistic applications and modeling.

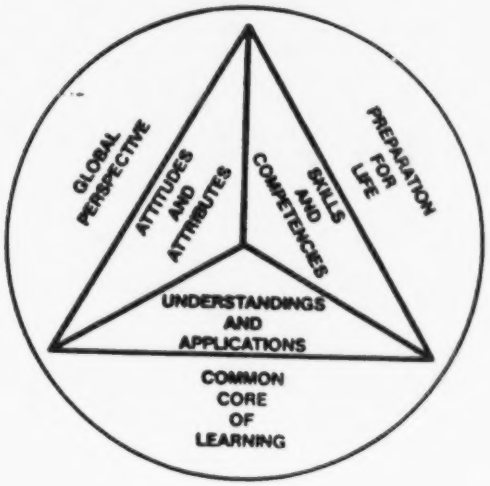
Connecticut's Strategy for Meeting the Critical Needs

The broader integrative paradigms called for in these national reports are incorporated into the proposed CCL Assessment Project design. Connecticut's approach addresses the desperate need for assessment (and ultimately instructional) models in science and mathematics education which require students to solve integrated and sustained multi-step problems and effectively use basic scientific inquiry and computation skills. The Connecticut assessment model is graphically displayed in Figure 1. The circle depicts the Department's commitment to a common set of skills, knowledge and attributes essential to the total development of all Connecticut students. The integration of communication and problem-solving skills with disciplines and attitudes is designed to directly address our concerns over NAEP's 1984 findings that "only 5% of our seventeen-year-olds can synthesize and learn from specialized reading materials and only 39% can find, understand, summarize and explain relatively complicated information." CCL tasks will necessitate synthesis, summary and explanation in the course of multi-step problem solving. The integration of these components in the CCL is not meant to define minimum competencies. Rather, it is a statement of Connecticut's high expectations for a broad range of student outcomes resulting from the K-12 school experience.

The cube in Figure 2 emphasizes the belief that learning is facilitated when: the learning experience is active rather than passive; the performance called for is more holistic than atomistic; and the thought processes required are divergent as well as convergent. Connecticut is setting out to create a CCL assessment project that is as broad and as multi-dimensional as the Common Core of Learning itself. The expectations are that the assessment project envisioned will provide a model of the kinds of authentic and integrated tasks that teachers can use for instruction as well as assessment. We believe that the time has come to develop assessments that are catalysts for the kind of learning that we value. The model currently in place -- atomistic tasks, passive learning and primarily convergent thinking -- has been too well served for too many years by multiple-choice testing.

Figure 1

Connecticut's Common Core of Learning Model



Attributes and Attitudes

Self-Concept
Motivation and Persistence
Responsibility and Self-Reliance
Intellectual Curiosity
Interpersonal Relations
Sense of Community
Moral and Ethical values

Skills and Competencies

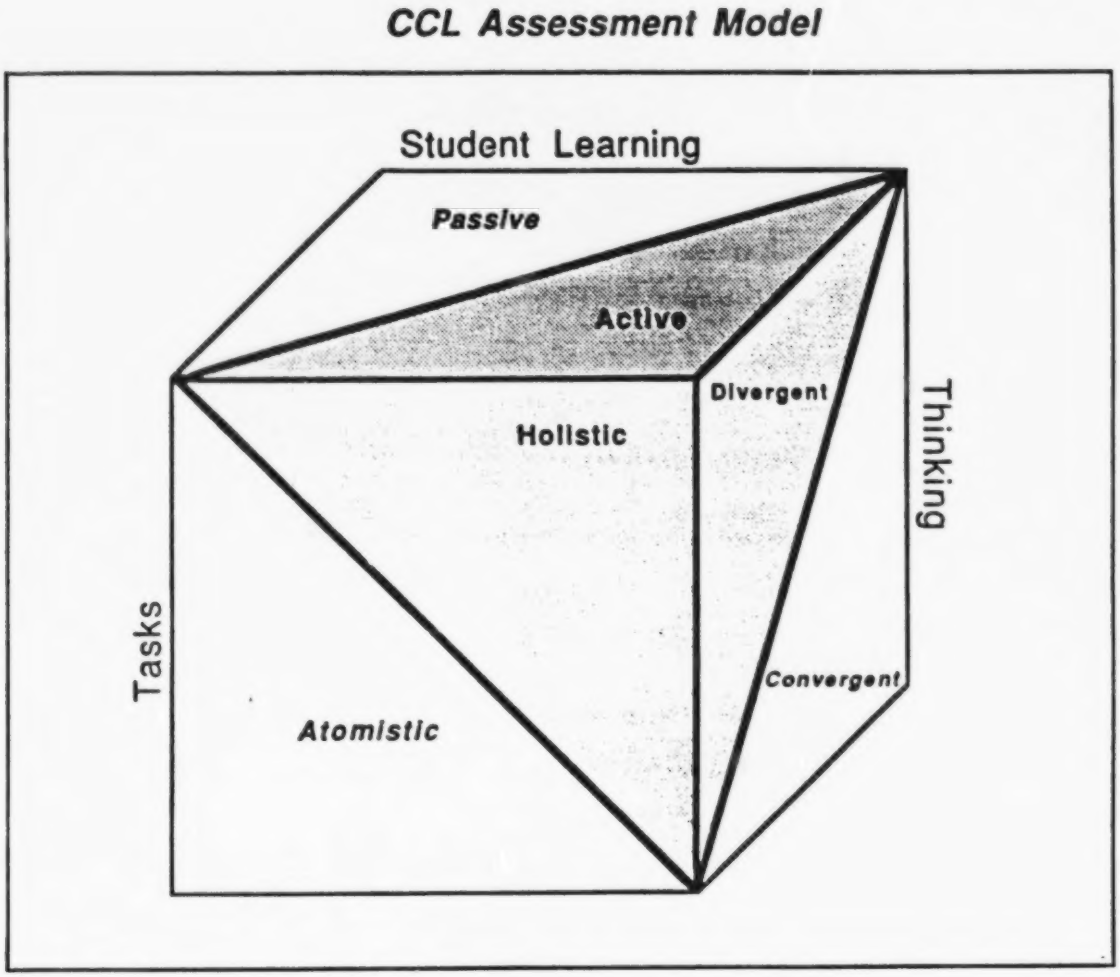
Reading
Writing
Speaking, Listening and Viewing
Quantitative Skills
Reasoning and Problem Solving
Learning Skills

Understandings and Applications

The Arts
Careers and Vocations
Cultures and Languages
History and Social Sciences
Literature
Mathematics
Physical Development and Health
Science and Technology

Figure 2

Connecticut's Common Core of Learning Assessment Model



C. Connecticut's Capacity and Commitment to Developing Performance Testing in Science and Mathematics

Given the current inadequate state of performance assessment methods and the need for better measures of student outcomes referenced above, Connecticut is particularly well qualified and prepared by its prior experiences in innovative assessment work and its demonstrated commitment to challenging conceptions of scientific disciplines to take the next step toward developing authentic and integrated performance tasks in the areas of science and mathematics.

Connecticut's Prior Experiences in Performance Testing

The Connecticut State Department of Education has built a strong foundation of performance testing through its Connecticut Assessment of Educational Progress (CAEP) Program and Connecticut Mastery Test (CMT) Program. Since 1980, the CAEP Program has developed performance testing in eleven areas (see Figure 3). Among these performance-based assessments was Connecticut's second science assessment conducted in 1984-85 (see Connecticut State Department of Education, 1986). This assessment included the use of science apparatus such as a triple-beam balance, graduated cylinders, a microscope and an electrical circuit. Furthermore, it required the design and execution of a science experiment modeled on tasks developed by the Assessment Performance Unit in Great Britain. This experience confirmed the feasibility of using performance tasks on future large-scale state assessment programs and gave us courage to consider some alternatives for more authentic and more sustained kinds of tasks.

The Connecticut Mastery Tests also include performance testing to assess the state's approximately 105,000 (total of students in all three grades) students at grades 4, 6 and 8 (see Figure 4). Students in all three grades produce a direct writing sample. As part of the CMT language arts test, students are required to take notes and use them to answer listening comprehension questions in response to tape-recorded messages. On the mathematics test, all eighth graders are required to use calculators on one portion of the mathematics test, which assesses realistic problem-solving applications that would otherwise be too time-consuming to include on a statewide test.

Developing New Indicators of Science and Mathematics Achievement through Performance Testing

Recent educational journals and conferences have been replete with articles about the pros and cons of measurement-driven instruction. Apart from whether this is a desirable condition, it is often true that high stakes state tests do have an impact on instruction (see Tirozzi, Baron, Forgione, and Rindone, 1985). Murnane (1987) has sensitized the educational community to the possibility of educational indicators being corrupted. Forgione and Baron (1987) and Baron and Forgione (1989) have described the catalytic role of educational indicators and the importance of developing indicators that capture the extent to which students report that they have internalized and practiced the behavior of experts.

Figure 3

Performance Testing in the Connecticut Assessment of Educational Progress Program 1980–87

Subject	Year	Grades Tested	Performance Task	Whole Sample or Subsample	Administration Time	When scored? (After self-administered testing or during other-administered testing)
Art	1980-81	4, 8, 11	Draw a room wall and draw a table with people around it	Subsample	1 class period	After
Music	1980-81	4, 8, 11	Sing "America" and complete a musical phrase	Subsample	A few minutes	During
Business and Office Education	1983-84					
Accounting		12	Make journal entries and complete a payroll record	Whole	1 class period	After
General Office		12	Timed typing	Whole	1 class period	After
Secretary		12	Type and compose part of a letter	Whole	1 class period	After
			Take shorthand	Whole	Part of a class period	After
English Language Arts	1983-84	4, 8, 11	Write 2 essays	Subsample	1 class period	After
			Take a dictated spelling and word usage exercise	Subsample	Part of a class period	
			Revise errors in focus, organization, support and mechanics	Subsample	1 class period	After
			Take notes from a taped lecture	Subsample	Part of a class period	After

continued

Figure 3 (continued)

Performance Testing in the Connecticut Assessment of Educational Progress Program 1980-87

continued

Subject	Year	Grades Tested	Performance Task	Whole Sample or Subsample	Administration Time	When scored? (After self-administered testing or during other-administered testing)
Science	1984-85	4, 8, 11	Use scientific apparatus: weigh, measure, focus microscope, etc.	Subsample	1 class period	During
		8, 11	Design and conduct an experiment	Subsample	1 class period	During
Foreign Language	1986-87	9-12	Write a letter	Whole	1 class period	After
French German Italian Spanish			Speak to an interviewer	Subsample	1 class period	During
Industrial Arts and Technology Education	1986-87					
Drafting		12	Produce a series of drawings	Subsample	3¼ hours	During
Graphic Arts		12	Produce a brochure	Subsample	5½ hours	During
Small Engines		12	Service and repair small engines	Subsample	3¼ hours	During

For information, contact Joan Boykoff Baron, Connecticut State Department of Education, P.O. Box 2219, Room 342, Hartford, CT 06145. (566-3847) (203)

Figure 4

Performance Testing in the
Connecticut Mastery Test Program
1985-88

Performance Testing in the Connecticut Mastery Test Program, 1985-88

Subject	Year	Grades Tested	Performance Task	Whole Sample or Subsample	Administration Time	When Scored? (After self-administered testing or during other-administered testing)
Language Arts	1985-1988	4, 6, 8	Writing Sample	Whole	45 Minutes	After
	1985-1988	6, 8	Take notes from a taped lecture	Whole	Part of a class period	After
Mathematics	1985-1988	8	Use calculator for problem solving applications	Whole	Part of a class period	After

Connecticut's Commitment to Innovative Performance Assessments

Connecticut's Common Core of Learning (CCL) assessment project builds upon Connecticut's earlier assessment efforts, under the CAEP and CMT Programs. In March 1986, Connecticut's Commissioner of Education, Gerald N. Tirozzi, convened a blue-ribbon committee of representatives from education, business and industry and the public-at-large to determine what their expectations were for Connecticut's students at the completion of high school. Their work focused on essential attitudes and attributes, skills and competencies, and knowledge and understandings. Drafts of their responses were circulated to thousands of educators over a year-long period. In January 1987, the Connecticut State Board of Education adopted Connecticut's Common Core of Learning.

Connecticut is committed to developing authentic performance tasks in all areas of the curriculum in conjunction with its assessment of Connecticut's Common Core of Learning. This proposal represents our first effort to launch the CCL and science and mathematics were selected to be the first areas to be assessed. In November 1988, the Department released a request for services (RFS) calling for experts in the areas of science, mathematics, measurement, and psychology to submit proposals. Thirty-eight proposals representing more than 50 individuals were sent in response, including a number of nationally prominent educators, psychologists, measurement specialists, and sociologists. Following a review process which included thirteen members of the Department and representatives from Connecticut school districts, several of these individuals were judged to be highly qualified and appropriate for working on the CCL assessment in science and mathematics. From this acceptable list, the Department identified a subset of national experts (see Figure 5) who will begin working with the Department under contract during the spring of 1989 to perform different combinations of the following tasks: define essential understandings in science and mathematics; conduct reviews of relevant literature; conceptualize the assessment framework; develop performance tasks and scoring rubrics; and identify measurement issues related to performance testing. The CCL developmental assessment activities occurring in Year 1 are entirely funded by the Connecticut State Department of Education.

Preliminary Blueprint for Prototype Performance Assessments

Although it should be noted that Connecticut is committed to developing authentic performance tasks in all areas of the curriculum in conjunction with its assessment of Connecticut's Common Core of Learning, the current project is for the development of assessments in science and mathematics only. The main components of Connecticut's development work are: task ingredients; prompt statements; task dimensions; scoring paradigm; rating system properties; identifying performance dimensions to be scored; evidence to be considered; and provisions for student feedback.

Figure 5

Common Core of Learning Assessment Development Team:
National Consultants

Dr. Elliot Aronson, University of California, Santa Cruz (UCSC)

Dr. Deborah Ball, Michigan State University

Dr. Edward A. Chittenden, Educational Testing Service (ETS)

Dr. Allan Collins, Bolt, Beranek & Newman (BBN) Systems and Technologies Corporation

Dr. Harris Cooper, University of Missouri-Columbia

Dr. Howard Gardner, Harvard University

Ms. Clare Heidema, Mid-Continent Regional Educational Laboratory

Dr. Ellen Langer, Harvard University

Dr. Barbara McCombs, Mid-Continent Regional Educational Laboratory

Dr. Jason Millman, Cornell University

Dr. Richard Stiggins, Northwest Regional Educational Laboratory

Dr. Hoi Suen, University of Connecticut

Dr. Leona C. Truchan, Alverno College

Dr. Joseph Walters, Harvard University

Dr. Grant Wiggins, Coalition of Essential Schools and CLASS Design and Management of Educational Change

. . . . . .

These national consultants will perform different combinations of the following tasks:

- o define the essential understandings in science and mathematics;
- o conduct reviews of relevant literature;
- o conceptualize the assessment framework;
- o develop performance tasks and scoring rubrics;
- o identify measurement issues related to performance testing.

In developing rich and sustained performance tasks, the CSDE will need to draw upon the expertise of nationally esteemed scientists and science educators, mathematicians and mathematics educators, measurement experts, social psychologists, and policymakers. Furthermore, six other states (e.g., Michigan, Minnesota, New York, Texas, Vermont and Wisconsin) and the Coalition of Essential Schools (Theodore Sizer and Grant Wiggins) have expressed interest in working with Connecticut in developing and pilot testing performance prototype tasks and rubrics.

In developing authentic performance tasks, Connecticut is committed to using real-world contexts to the extent possible. This commitment is grounded on many different lines of research. For example, the tasks that are developed will attempt to extend into the assessment domain recent work in "anchored instruction" by Bransford, Sherwood, Hasselbring, Kinzer, and Williams (in press), and "situated cognition and the culture of learning" by Brown, Collins, and Duguid (1989) which build upon earlier work reported by Rogoff and Lave (1984). Our assessments are also concerned with issues of "transfer" (see Perkins and Salomon, 1989) and will directly address Whitehead's (1929) distinction between "inert knowledge" and the development of "useful, robust knowledge." Brown, et.al. (1989) note that "it is common for students to acquire algorithms, routines, and decontextualized definitions that they cannot use and that, therefore, lie inert. Unfortunately, this problem is not always apparent."

The assessment instruments designed for this project will directly address this concern. Our goal is to develop integrated assessment models which call upon students to use mathematical algorithms and scientific concepts within practical real-world contexts. To the extent that we succeed in doing so, there is the likelihood that curricular and instructional changes will take place.

Collaboration with the Coalition for Essential Schools

One major influence on our conceptualization of this CCL assessment has been the work of Theodore Sizer and Grant Wiggins, done in conjunction with the Coalition of Essential Schools. We are quite pleased that these two individuals will be assisting us. In his book, Horace's Compromise, Sizer suggested the use of student exhibitions as an alternative assessment model. Although several schools have embraced the Coalition concept of student-as-worker, only a small handful of them have actually implemented student exhibitions as a viable assessment model. One of these schools, Walden III, is in Racine, Wisconsin. In October 1989, Dr. Joan Baron (Program Director, CAEP Program) spent a full day at Walden III meeting with Charles Kent, the principal, Tom Feeney, the teacher of the senior class preparing for the Rights of Passage Experience (ROPE) and other teachers and students in the school. Connecticut hopes to benefit from the experiences of Walden III and other high schools which are part of the Coalition.

D. Major Common Core of Learning Project Tasks and Activities

Over the next three years, the CSDE proposes to address the following major tasks . They include:

Task 1: defining the essential understandings in science and mathematics;
Task 2: selecting a statewide sample and soliciting participation of local option schools;
Task 3: planning and conducting the Summer Institutes for Teaching and Learning (ITL) workshops;
Task 4: identifying psychometric issues and considerations related to the validity, reliability and practicality of these instruments;
Task 5: implementing the performance tasks and rubrics;
Task 6: reporting on the condition of science and mathematics education; and
Task 7: releasing performance tasks and scoring rubrics for use in classrooms in Connecticut and in the Nation.

These tasks are described below in this section.

Task 1: Defining the Essential Understandings in Science and Mathematics

In developing the framework for the integrative tasks to be developed in science and mathematics, it is critical to determine those essential understandings in each domain; that is, what do scientists and mathematicians really want students to be able to do with mathematics and science by the end of high school? This addresses the "big picture" issues in science and mathematics and is intended as a safeguard to ensure that the forest isn't mistaken for a simple array of trees. There are several ways to conceptualize this. In Everybody Counts, the authors refer to some overarching mathematical understandings which students apply across the various content areas. Similarly, frameworks in science have attempted to identify certain concepts that cut across all domains of science (e.g., change, symmetry) whereas others have looked instead for concepts that unify a single science domain. Another approach has been taken by the authors of the 2061 Report who refer to the outcomes of learning as the "residuals", i.e., "the knowledge, insights and skills that people should possess after the details have faded from memory" (p. 18, October 1988 draft). However, the 2061 Report authors would undoubtedly be quick to point out that not all residuals are of equal importance. Therefore, the CCL project advisors will need to identify and specify the project priorities early in the developmental effort.

Without such a framework, the development of tasks could default into a reflection of the interests and expertise of the developers and neglect other areas which are of less interest or are more difficult to measure. This latter group of objectives that are difficult to measure will have a tendency to fall by the wayside unless a framework exists which ensures that the essential understandings have a serious opportunity to be explored. This will be a subject of consideration and serious deliberation for early meetings of both the National Advisory Committee and the Statewide Advisory Committee as these decisions will guide all future development activities. In this activity, we will benefit from the work of other states like Michigan in Science and Vermont in Mathematics. Both of these states have begun to

address these questions as they develop frameworks for their own new assessments. It should be noted that a parallel set of activities will take place among the social and educational psychologists in defining the most essential manifestations of the attitudes and attributes included in the CCL.

Task 2: Selecting a Statewide Sample and Soliciting Participation in the Connecticut Local Option

The CSDE will select 45 Connecticut high schools out of some 130 to participate in the state sample. Fifteen schools will be chosen for science only, fifteen schools will be chosen for mathematics only, and fifteen schools will be chosen for both science and mathematics. As part of our commitment to the importance of training teachers in the development of alternative strategies, the Department will offer a local option as part of this CCL assessment. Any district which is not included as part of the forty-five school statewide sample will be invited to join the assessment and implement the performance tests in its schools. The results from the local option schools will not be included in the statewide sample but will be available for use by the participating schools. Training will be provided by the Department in the administration and scoring of the tasks; however, local option districts will be expected to assume the cost of scoring their tests.

Task 3: Planning and Conducting the Summer Institutes for Teaching and Learning (ITL) Workshops

During Year 2, which begins in the summer of 1989, the Department will train 25 Connecticut mathematics teachers and 25 Connecticut science teachers through summer institutes held in conjunction with Connecticut's Institutes for Teaching and Learning (ITL). In addition, science and mathematics teachers from Michigan, Minnesota, New York, Texas, Vermont and Wisconsin (and one or two other states) and faculty from schools participating in Theodore Sizer's Coalition of Essential Schools will attend the summer training sessions. The pilot tasks developed in Year 1, as well as additional tasks and scoring rubrics developed by these teachers during the summer and fall, will be tried out in their classrooms during the 1989-90 school year. As areas of need are identified, these teachers will assist the project by developing additional tasks.

The Department will provide follow-up seminars during the 1989-90 school year at which the 50 Connecticut science and mathematics teachers will meet bi-monthly and communicate through teleconferencing and computer networks with the out-of-state teachers. Several national and Connecticut curriculum, instruction and assessment experts will assist this group of teachers to refine their new tasks for possible implementation in the CCL assessment in the following year. The Connecticut teachers will become resources to the CCL project to assist in the training of other teachers both within their schools and in other schools during Year 3. Field supervisors and will visit some, or all, of the Connecticut teachers participating in the summer workshops. The teachers from other states and the Coalition will become resources and demonstration sites within their own states for the viability of developing and implementing performance assessments in high school science and mathematics classrooms. A second round of ITL workshops will be conducted in the summer of 1990 for Connecticut science and mathematics teachers.

Task 4: Identifying Psychometric Issues and Considerations

A group of psychometric experts will continue to assist Department staff in an ongoing effort to establish the validity and reliability of the results and develop a strategy for aggregating the results from the separate tasks into larger reporting units to conform to the categories delineated in the CCL. In this way, the CCL staff will be able to report to the Connecticut State Board of Education in the fall of 1991 on: the performance of Connecticut's high school students in the knowledge and understanding of science and mathematics; the integration of science and mathematics with the communication skills of reading, writing, speaking, listening and viewing; the appropriate use of reasoning, problem-solving and learning skills; and the manifestation of attitudes and attributes such as positive self-concept, motivation and self-reliance elicited by the science and mathematics performance tasks.

Issues of Reliability

An important feature of any performance assessment instrument is that the scores it yields are as free as possible from errors of measurement. Measurement errors are imperfections which prevent one from assessing the correct levels of students' current abilities. Measurement errors can be inferred from differences between two observers rating the same tasks; differences between scores on tasks which have been designed to measure the same traits; or differences in scores from one occasion to the next which cannot be attributed to learning or maturation.

The use of raters to observe and score student performance on CCL tasks has an inherent risk of low reliability unless: (a) the behaviors identified for observation and rating are inherently clear and stable; (b) rating procedures are carefully constructed; and (c) raters are thoroughly trained and monitored in their use. A major charge of the Psychometrics Committee will be to assure that the procedures designed for developing, administering and scoring tasks meet these conditions.

The reliability of all tasks and scoring procedures used will be monitored over the course of this project. In particular, we will be attempting to answer the following questions about this genre of assessment tools:

a) **Inter-rater reliability.** Will two raters observing performance on the same CCL task tend to produce similar total scores and dimension scores for students? If they do not, can the problem be localized to rating scale flaws, rater flaws or problems concerned with the nature of the tasks themselves?

b) **Cross-Problem Dimensional Reliability.** Do ratings based on CCL attributes (i.e., cooperativeness, motivation, persistence) show some degree of consistency across tasks--or do these ratings greatly fluctuate? If they fluctuate, is this due to errors in the measurement system; to real differences such as actual improvement due to growth or learning; to differences in task demands; or to the way these dimensions interact with one another (e.g., does a student who knows less about a particular subject area also show lesser levels of motivation, group interaction, etc.)?

c) **True Score Reliability**. A basic question which will have an elusive answer in the real world is: would the student's performance on a task approximate his/her "true score" if this could be known? Put another way, how much change would occur in the typical student's score if he/she could take the same task again under the same conditions, without prior knowledge of the results? This question cannot be addressed directly since the performance tests would not be readministered. However, the identification and estimation of sources of measurement error may help to approximate some answers to this important question.

The issues of reliability to be addressed are clearly more complex than those encountered with traditional, objectively scored multiple-choice tests where reliability has been handled with split-half correlations and alpha internal consistency indices. Our Psychometrics Committee, many of whom are nationally recognized measurement experts, will offer guidance in this area.

Issues of Validity

While reliability is an important feature to consider, validity is the single most important feature for any assessment procedure to achieve. Whereas reliability data address the concern of whether scores are consistently drawn, validity data address the question of how meaningful those scores are, i.e., whether they actually reflect what they set out to measure.

There is currently a lack of adequate evidence of empirical validity for performance tasks of the type contemplated in this project. Although there is much positive to be said about the "ecological validity" or faithfulness to a real-world context, more concrete demonstrations of validity will be a major project goal.

As with the issue of reliability, the issue of validity has two aspects to consider: building features into the performance tasks and rating scales which will foster validity; and monitoring whether this process has actually been achieved. The first of these aspects primarily addresses issues of content validity, while the second addresses construct and criterion-related validity.

Content and Ecological Validity. To achieve content validity, the performance assessments must relate directly to important instructional outcomes. For our purposes, this means that performance tasks must be closely tied to the goals set forth in the Common Core of Learning. In addition, some tasks should be "authentic" and relate to tasks that students would be expected to perform in the world outside of school.

A first step taken to build in this feature is the development of a draft set of "Guidelines for the Development and Use of Statewide Common Core of Learning Assessments" (see Section II). These guidelines are intended to focus developers on issues which will maximize the relevance of the tasks developed; identify the behaviors to be focused upon in scoring; and guide the way those behaviors will be evaluated and categorized. An additional project component will be the establishment of a committee review process for all task prompts and scoring systems developed, to ensure the inclusion of the essential questions of the discipline, and that the tasks conform to the standards set forth in the "Guidelines.". These standards, embodied in part in the "Guidelines" document, will be refined with the assistance of the Psychometric Committee.

Construct and Criterion-Related Validity. Building relevance into the CCL performance tasks, and demonstrating that this has been done, are two important components in establishing their validity. We also wish to explore the question of validity in a more empirical sense, by addressing the issue of how the scores obtained on CCL tasks relate to other real-world behaviors which are considered to be related to competence in the areas tapped by the assessments. Here, a perennial problem in validation research arises: what is the appropriate criterion measure or measures against which to compare scores to assess their validity? There are numerous indicators of competence or success which may be considered (for example, course grades aside from CCL task scores; scores on standardized tests; scores in other classes; self-ratings; other achievement measures) but each has its own limitations.

The design of an appropriate validation strategy will be a complex task. Several key questions to be addressed are:

- Do performance tasks measure basically the same things as are already reflected in traditional testing and grading practices?
- If these two methods differ, then how do they differ?
- Do performance tasks show more evidence of "validity" or relationships to important real-world indicators of competence than do traditional testing and grading procedures?
- Are there components underlying performance on CCL tasks which are generally unmeasured by traditional tests (e.g., cooperativeness, persistence) which can be identified as making a valuable demonstrably valid contribution to task performance if students can properly apply or otherwise manifest them?

A comprehensive strategy to address these broad questions will be designed with the input of the Psychometric committee. In so doing, we will be guided by the American Psychological Association's Standards for Educational and Psychological Testing (1985), as well as by the thinking of distinguished authors on the topic of validating complex tests (e.g., Messick, 1986, Cronbach, 1986).

Issues of Feasibility/Practicality

While attempting to maximize the reliability and validity of the performance tasks developed, we must also be equally mindful of the issues of feasibility and practicality. Assessments which have proven high performance from a purely psychometric standpoint will not fulfill their purpose, and will not be used, if they are perceived to be highly complex and time-consuming, or if the results cannot be easily interpreted. Furthermore, many of these tasks will provide two levels of scores -- one for each participating student and another for small groups of students working together to solve problems. The project team will therefore build in safeguard procedures in the development, review and implementation phases to assure that, aside from their technical soundness, these tasks and exercises will:

- Minimize the "teacher burden" required to administer them. Ideally this would be accomplished if the tasks are instructionally relevant.

- Minimize the scoring burden, through such measures as manageable numbers of dimensions to be rated, clear indications of the behaviors relevant to the dimensions and how they would be evidenced in performance, and clearly conceptualized rating categories.

- Be understandable to students by giving them meaningful feedback on their performance and allowing them to use the scoring rubrics to guide and assess their own performance.

- Be capable of aggregation into meaningful summaries which can be reported to the Connecticut State Board of Education, school officials and the public in the form of a "report card".

Task 5: Implementing the Performance Tasks and Rubrics

In the third year, 1990-91, the Department will implement the first series of performance tasks in Connecticut. The tasks will be scored by both teachers and students within the participating schools and by an independent team of scorers from outside the school. This will address a need for objectivity and the demonstration of acceptable levels of inter-rater reliability. Preliminary specifications for the scoring rubrics are provided in Section II. The science and mathematics associates will play a key role in visiting all of the schools participating in the statewide sample. They will help to ensure a sufficient level of training for teachers and students in the application of the scoring rubrics. They will report any difficulties or irregularities in the implementation of the prompts and exercises. Should any last minute modifications be necessary, this group will be the first line of communication.

Task 6: Reporting on the Condition of Science and Mathematics Education

The most immediate use of the tasks and rubrics developed for this state assessment is for reporting on the science and mathematics performance of Connecticut's high school students to the Connecticut State Board of Education, which will occur early in Year 4 during the Fall 1991.

The science and mathematics teaching forces in Connecticut and nationally are our second audience. There have been many nationally disseminated studies which lament the fact that teachers rely almost entirely on multiple-choice tests and the recall of specific knowledge for assessing and grading their students. Consistent with this emphasis on recall is a heavy reliance on teacher-directed activities where teachers dispense information to passive learners. One goal of this assessment is to implement activities which assume students as more active learners who will help to shape their own questions and strategies. Such a model requires teachers who view themselves as facilitators of the learning process.

Task 7: Releasing Performance Tasks, Rubrics and Handbooks for Use in Classrooms in Connecticut and in the Nation

In year 4, the CSDE will release some, or all, of its CCL performance tasks and scoring rubrics for teachers to use in their classrooms. There will also be accompanying handbooks describing strategies for developing and scoring other performance tasks. Our hope is that by providing valid and reliable models of performance tasks for use in classrooms, teachers will be able to directly observe their advantages, and will move from an over-reliance on multiple-choice testing to the implementation of these more generative and productive assessment strategies.

In the pages which follow in Section E. is a draft of guidelines that are currently being used to develop sample prompts in mathematics and science.

E. Guidelines for Development and Use of Statewide Common Core of Learning Assessments in Grades 9-12

I. Rationale for Common Core of Learning (CCL) Assessment

The Common Core of Learning (CCL) is a statement of the student outcomes expected to result from the entire K-12 school experience. It has been developed to influence curriculum by generating discussion and stimulating change in school programs, student objectives, resource allocations and teaching. It articulates the expectations held for Connecticut's high school graduates. It was clear from the outset of CCL development that the vision, breadth and depth of the CCL would require a developmental effort to construct new assessment instruments which focus on students' abilities to demonstrate "knowledge-in-use." Traditional measurement methods, such as multiple-choice tests that focus on convergent thinking, would constrain and narrow the CCL vision. Through new forms of assessment, we hope to demonstrate the value of the "student-as-worker" who is actively engaged and expected to construct, investigate, translate, discuss, analyze, and formulate in a dynamic way. The CCL is a vision of a student whose learning is not only a matter of retaining information but of constructing understanding and whose inquiry mode is self-directed.

This development effort also has important implications for: (1) the integration of curriculum, instruction and evaluation; (2) a new genre of teaching and classroom assessment methods such as exhibitions, 'hands-on' performances, field and laboratory experience, portfolios, simulations, projects, and oral and written communications; and (3) the development of divergent and higher order thinking in the classroom. This, of course, means that teachers must develop alternative methods which involve students in a more active role in their own education.

A brief overview of the CCL assessment program is presented below.

COMMON CORE OF LEARNING ASSESSMENT OVERVIEW

PRIMARY PURPOSE — To assess the vision, breadth and depth of the Common Core of Learning (CCL) by providing intellectual challenges which require students to demonstrate "knowledge-in-use." These results will be reported to the State Board of Education.

DESIGN — An assortment of assessment strategies will be designed which are largely performance-based (exhibits, portfolios, projects, reports, group discussions, speaking and writing assignments).

CONTENT — Performance tasks will be designed to measure the integration of students' academic knowledge, skills and attitudes essential to the total development of all students. Performances will be oriented around a set of essential concepts that address the "heart" of a particular discipline.

LOCAL DISTRICT USES — The CCL assessment will provide models for local districts to use to aid curriculum and instruction by:

- providing baseline data for monitoring changes over time;
- serving as a model for the integration of curriculum, instruction and evaluation;
- training teachers, statewide, in the understanding and rating of performance tasks; and
- providing students and teachers with an understanding of the CCL's expectations and the criteria of good performance in advance of the administration of the assessment.

A brief overview of the CCL assessment program is presented below:

- Tasks. Tasks and scoring should be less atomistic (e.g., addition of two-digit numbers with regrouping) and more holistic (e.g., sustained thinking, integration, and application of the basic principles, concepts and language of a discipline to an issue or problem) in nature. Students should be required to apply, synthesize and integrate understandings in a discipline rather than simply recall and/or repeat memorized content.

- Thinking. Students should be required to demonstrate and engage in more divergent thinking in the accomplishment of performance tasks. Traditional measurement methods, such as multiple-choice, tend to emphasize convergent thinking.

- Learning. Active learning implies the idea of the "student-as-worker" and the "teacher-as-coach." The CCL assessment will measure students' abilities to demonstrate "knowledge-in-use" by requiring them to construct, investigate, translate, discuss, analyze, formulate and apply.

II. Specifications for Performance Tasks

The following specifications (Sections II and III) will govern the development of CCL assessment performance tasks.

A. Task "Ingredients"

The following are the components, or "ingredients" that must be considered in the design of mathematics and science performance tasks. These "ingredients" comprise the three major dimensions of the CCL.

1. Attitudes and Attributes

Prompts must require students to perform or respond, individually and/or in groups, in ways that make it possible for classroom teachers to observe (in some cases, unobtrusively) and rate student attitudes and attributes as demonstrated in the performance of science and mathematics tasks. Assessing attributes such as persistence, motivation, curiosity, ethical values and effective interpersonal relations (see Attachment A, pages 5-7 for an expanded list) in relation to performance tasks is of critical concern in an integrated assessment.

2. Skills and Competencies

a. Critical Thinking

Prompts must contain verbs or action words that elicit some higher order thinking (e.g., use concepts and generalizations, identify and formulate problems, develop solutions, make predictions and hypotheses based on data, etc.) See page 10 of Attachment A for an expanded list. Performance tasks which elicit only recall and comprehension of content will not be acceptable.

b. Communication Skills

Prompts must require students to exhibit a variety of communication skills, such as, reading, writing, speaking, listening, viewing and quantifying (see Attachment A, pages 8-10 for an expanded list).

3. Understandings and Applications

Prompts must include opportunities for students to apply their knowledge and understanding of mathematics and science (see pages 15 & 16 of Attachment A for an expanded list) through tasks, organized around key understandings which (1) go to the "heart" of a discipline, i.e., address important concepts/aspects of a discipline, and (2) may have multiple correct answers. It is important that, through performance tasks, students and teachers understand that the major purpose of acquiring content knowledge is to answer significant questions and solve relevant problems. The prompt should also require opportunities for sustained thinking on an issue and the integration of the basic principles, concepts and language of a discipline. For example, geometry prompts should, to some extent, require students to demonstrate an integrative understanding of general mathematics, algebra and geometry.

II. Specifications for Performance Tasks (continued)

B. Prompt Statements

Prompts should contain, when feasible, four types of statements (Task, Elaboration, Strategy and Focus) to help students understand the task before them.

1. Task Statement

The Task Statement directs the student to perform one or more tasks on a specific topic, described by key words or phrases organized around a theme. This theme, or the student's narrowing of it, serves as the focal point of the student's performance. Task Statements begin with words like "Write about...", "Display the...", "Record and classify the...", "Design a...", and conclude with the theme phrase.

2. Elaboration Statement

The Elaboration Statement provides examples or further definition of the theme. The intent is to provide a common understanding of the meaning of the theme by expanding, restating, or clarifying the central idea for the student. It is not the intent to preclude the student's narrowing or restating of the theme to suit his or her own plan.

3. Strategy Statement

The Strategy Statement is intended to broadly suggest approaches to the topic to help students get started in their project planning. The purpose of the Strategy Statement is to lend some focus to student planning efforts, but <u>not</u> to provide the specific steps of the plan, which should be produced by the student. An example of a strategy statement is "You might want to spend part of your planning time thinking about..." and/or some brief examples of alternative ways of looking at the problem.

4. Focus Statement

The Focus Statement should help the student accumulate the information or content that ought to be delivered. It is also a reminder of the student's responsibility to the prospective audience. The Focus Statement may begin with the phrase, "Think about information that will help your audience understand...".

C. Desirable Task Characteristics

1. Ecological Validity

Prompts should present situations which closely resemble (1) the "real world" of high school students, and (2) the requirements of a technological society. That is, tasks should reflect what we expect students to be able to do with knowledge learned as a result of 180 days of biology, algebra, etc. Stated differently, "Can students apply what they have learned to their own lives?".

II. Specifications for Performance Tasks (continued)

2. Practicality

Tasks should be designed to minimize the "teacher burden" required to administer them. In other words, performance tasks should be valuable learning activities which do not require unrealistic amounts of teacher time or support.

Tasks should minimize the "scoring burden", through such measures as manageable numbers of dimensions to be rated, clear indications of the behaviors relevant to the dimensions and how they would be evidenced in performance, and clearly conceptualized rating categories.

C. Task Dimensions (continued)

Tasks should be understandable to students by giving them meaningful feedback on their performance and allowing them to use the score system to guide them in understanding and improving their performance.

3. Check Points

Prompts should include the identification of key stages/check points so that teachers can periodically assess and record pupil progress. This will permit a comprehensive view of a pupil's ability and knowledge to perform a given task and to determine which students need additional assistance.

D. Task Categorization

CCL tasks can be usefully categorized along two primary dimensions: task length and task participation (individual or group). These two dimensions are described below.

1. Task Length

Three types of tasks, each requiring different lengths of time to complete, may be elicited by prompts.

a. Short Tasks

Short tasks may take part of a class period to complete, possibly with some additional preparatory homework. In performing these tasks students would be demonstrating mastery of a highly limited number of skills/competencies, attitudes/attributes and understandings/ applications.

b. Long Tasks

Long tasks may take the equivalent of several class periods with work time divided between in-class and out-of-class activities. In performing these tasks students would need to demonstrate mastery of (1) a greater variety of skills/competencies, attitudes/ attributes and understandings/applications and (2) an understanding of the overlap and integration of concepts in a discipline.

II. Specifications for Performance Tasks (continued)

c. Extended Tasks

Extended tasks may take as long as one semester. Work time will be divided between in-class and out-of-class activities. These are long term projects that require students to apply and integrate a wide range of skills and understandings and display attitudes/attributes conducive to good performance. In the completion of extended tasks, students should be required to keep logs of their progress. The organization of the log should be specified by the prompt developer. If properly completed, logs can be useful in rating processes which are not readily observable.

2. Task Participation

a. Individual

These tasks would be performed entirely by individual students.

b. Group

These tasks would be performed by small groups of students working cooperatively and sharing credit equally for the final product(s) of the task. In some cases, it is useful for students to work in groups in order to facilitate development of the maturity and skills required to perform individual tasks of high complexity. Just as importantly, group tasks give students the opportunity to develop cooperative and interactive behaviors in a work-oriented setting.

c. Mixed

These tasks would contain both individual and group elements. That is, some aspects of the task would be performed cooperatively and the group would share the evaluation equally, while other aspects of the task would be performed by the individual members of the group.

It is important that the Connecticut State Department of Education receive a variety of prompts that take length (short, long and extended), type of participation (individual, group or mixed) and into account. These dimensions are graphically depicted below.

		PARTICIPATION		
		A Individual	B Group	C Mixed
LENGTH				
A	Short			
B	Long			
C	Extended			

III. Specifications for Scoring Rubrics

Performance assessments tend to focus on complex real-world behaviors rather than structured, easily-scored responses. As such, they require a great deal of attention to scoring procedures. Attention to scoring issues should begin at the very outset of exercise development. The desired task outcomes and the dimensions of performance to be observed and rated should be considered and written down in the early stages of exercise development, so that the exercise has a clear purpose and clear assessment goals.

A. The Scoring Paradigm

The typical steps in developing and applying a scoring system for performance assessments is as follows:

1. The developer of the exercise specifies its purpose (e.g., student outcomes expected) and the dimensions to be rated. This must be done at the outset of exercise development (see III.C. below).

2. The dimensions to be rated are operationally defined within the context of the Common Core of Learning and linked to quality levels on a rating scale (see III.C).

3. The evidence to be considered in support of each dimension score is specified (see III.D.).

4. Once the exercise is administered and completed, the rating system is applied to all evidence which relates to the dimensions and levels of performance exhibited. A series of dimension scores is then arrived at for each student.

5. A final assessment profile is produced for each student, including ratings and feedback to students, to explain the reasons for the ratings and point the way to improved performance.

6. Scores are maintained in such a way that they are available by task, by dimension, and for groups as well as individuals.

B. Rating System Properties

The scoring system must be sufficiently structured to allow: (a) linkage to some pre-set standards of quality; (b) reasonable levels of agreement across raters; and (c) comparisons across students, who may be performing substantively different tasks or projects.

The scoring of performance assessments must necessarily rely upon judgmental ratings, and the exercise developer should be aware of how easily bias and judgmental errors can "creep" into such ratings. To minimize such rating errors, care must be taken to develop a rating system which is systematic, explicit (behavior-specific), and objective to the extent possible. This is best accomplished by a system which carefully defines the dimensions to be assessed; clearly defines the behaviors indicative of good and poor performance on a dimension; and provides for notes to further clarify and justify the ratings made.

III. Specifications for Scoring Rubrics (continued)

B. Rating System Properties (continued)

Whenever possible, more than one rater should be used. This is critical to demonstrating that the system is sufficiently structured and objectified, so that two raters observing the same student tend to arrive at similar ratings.

C. Identifying Performance Dimensions to be Scored

Just as task prompts must be designed to elicit performance along a wide variety of dimensions, so should the scoring system be designed to allow the observation and rating of these dimensions. To the extent possible, the scoring system should allow for the observation and rating of higher order processes, knowledge, skills, competencies, attitudes and attributes. The performance dimensions to be evaluated should be mapped against the Connecticut Common Core of Learning, which delineates desired student outcomes.

A performance dimension should be included in the scoring rubric only if it can be effectively observed in task performance or inferred from a task product, and reliably rated with a minimum of inference required.

In developing the scoring criteria, each performance dimension to be assessed must be operationally defined. In other words, the developer must specify what is meant by the performance dimension; how it is elicited by the exercise; and what types of behavior will be taken as evidence for the dimension. Each dimension should be accompanied by a rating scale which includes qualitative descriptions and behavioral examples corresponding to different levels of performance.

In scoring the results of a performance assessment, the rater(s) must consider all evidence which relates to the dimensions and their associated levels of magnitude. Such evidence should largely have been specified in advance, in the scoring rubric. Evidence might include one or more of the following:

- o Student portfolio, journal or log of the project;
- o Written reports
- o Oral reports
- o Group discussion (e.g., observers' notes)
- o Teacher observations (events, interventions, etc.)
- o Student self-assessments
- o Final product(s) of the task

In a performance assessment, products or performances can be rated. There are advantages and disadvantages to each approach. Since products are more concrete and lasting, they can be used as a stable point of reference when making ratings or discussing them. However, it is harder to infer attitudes and attributes from products. Performances tend to provide richer examples of attitudes and attributes such as motivation or cooperative behavior. However, performances do not create a permanent record (unless recording is feasible).

III. Specifications for Scoring Rubrics (continued)

D. Provisions for Student Feedback

The authentic nature of performance assessments makes feedback all the more important for students, to bring about performance improvement. To pave the way for this feedback, the scoring system should meet the following specifications:

1. In most cases, the evaluation criteria should be communicated to the students when they are given their assignments, so that they know in advance how their performance will be evaluated, and are better able to monitor their own progress.

2. The scoring approach used should allow for both a specific rating corresponding to a level on the rating scale, and narrative comments to more specifically describe performance strengths and weaknesses.

3. Exercise performance should be summarized on a rating sheet which includes the dimensions to be rated; the levels of quality associated with each dimension; the dimension ratings assigned; and appropriate narrative justification for each rating.

IV. Early Development Work

Over the past several months, the Department staff have begun exploratory work on developing models or prototypes for illustrating the proposed authentic and integrated performance assessments in the area of mathematics. Obviously, there is much to do over the next nine months (i.e., prior to the first proposed pilot testing next winter/spring 1990), but our initial exploratory work does offer a preliminary glimpse of the type of developmental effort Connecticut is contemplating. Attachment 1 (pp. 31 ff) provides an example of a sample mathematics prompt.

The Connecticut State Department of Education is interested in maintaining ongoing communications and sharing new understandings with individuals and projects that share this challenging vision of the future of integrated performance assessment approaches in the areas of high school science and mathematics.

Bibliography of Pertinent Literature

Alverno College Faculty (1979). Assessment at Alverno College. Milwaukee, Wisconsin: Alverno College.

American Association for the Advancement of Science (1988). Science for all Americans: Literacy goals in science, mathematics and technology. Washington, D.C.

American Educational Research Association, American Psychological Association and National Council on Measurement in Education (1985). Standards for educational and psychological testing. Washington, D.C.: American Psychological Association.

Baron, J.B. and Forgione, P.D., Jr. (1989). Collecting and profiling school/ instructional variables as part of the State-NAEP results reporting: Some technical and policy issues. In E. Haertel (Editor and Chair). Report of the NAEP technical review panel on the 1986 reading anomaly, the accuracy of NAEP trends, and issues raised by state-level NAEP comparisions. National Center for Educational Statistics Technical Report. CS89-499 United States Department of Education (OERI). January.

Blank, R. (1989). Conference Organizer of: Conference on alternative methods of assessing science. Sponsored by the Council of Chief State School Officers and the National Science Foundation. Tampa, Florida. January 13.

Bransford, J.D., Sherwood, R.D., Hasselbring, T.S., Kinzer, C.K., and Williams, S.M., (in press). Anchored instruction: Why we need it and how technology can help. In D.Nix and R. Shapiro (Eds.) Hillsdale, N.J.: Erlbaum.

Brown, J.S., Collins, A., Duguid, P. (1989). Situated Cognition and the Culture of Learning. Educational Researcher. 18 (1) 32-42.

Commission on Standards for School Mathematics of the National Council of Teachers of Mathematics (1987). Curriculum and evaluation standards for school mathematics. (Working Draft). Reston, Virginia. National Council of Teachers of Mathematics.

Connecticut State Department of Education (1986). Connecticut Assessment of Educational Progress 1984-85 Science summary and interpretation report. Hartford, Connecticut.

Connecticut State Board of Education (1987). Connecticut's Common Core of Learning. Hartford, Connecticut. January.

Cronbach, L.J. (1980). Validity on parole: How can we go straight? in New directions for testing and measurement proceedings of the 1979 ETS Invitational Conference. (pp. 99-108). San Francisco: Jossey-Bass.

Dossey, J.A., Mullis, I.V.S., Lindquist, M.M., Chambers, D.L. (1988). The mathematics report card: are we measuring up? Princeton, N.J.: The Educational Testing Service.

Forgione, P.D., Jr. and Baron, J.B. (1987). Can reporting on educational indicators serve as a catalyst for the improvement of educational achievement? A visionary exploration. Presented at the annual meeting of the American Statistical Association. San Francisco, California.

Harmon, Mokros, Dawson, Hartwig, Henderson, Lowery, and Taylor (1988). Comments on the need for new assessment materials. Paper submitted to National Science Foundation. June 23. (4 pages)

Hein, G.H. (1988). Organizer of: An Invitational Elementary Science Assessment Conference. Sponsored by the National Science Foundation. Chatham, Massachusetts. November 4-6.

International Association for the Evaluation of Educational Achievement (IEA) (1988). Science achievement in seventeen countries. Oxford, England. Pergamon Press.

McKnight, C.C., Crosswhite, F.J., Dossey, J.A., Kifer, E., Swafford, J.O., Travers, K.J., and Cooney, T.J. (1987). The underachieving curriculum: assessing U.S. school mathematics from an international perspective. Champaign, Illinois. Stipes Publishing Company.

Messick, S. (1986). The once and future issues of validity: assessing the meaning and consequences of measurement. Educational Testing Service Research Report.

Mullis, I.V.S. and Jenkins, L.B. (1988). The science report card: elements of risk and recovery. Princeton, New Jersey. Educational Testing Service.

Murnane, R.J. (1987). Improving education indicators and economic indicators: The same problems? Presented at the annual meeting of the American Educational Research Association. Washington, D.C. April.

National Research Council (1989). Everybody counts: a report to the nation on the future of mathematics education. Washington, D.C. National Academy Press.

Pecheone, R., Baron, J.B., Forgione, P.D., Jr., and Abeles, S. (1988). A comprehensive approach to teacher assessment: Examples from math and science. In A. Champagne (Ed). This year in school science 1988. Washington, D.C. American Association for the Advancement of Science.

Perkins, D.N. and Solomon, G. (1989). Are cognitive skills context bound? Educational Researcher 18, (1) 16-25.

Raizen, S.A., Baron, J.B., Champagne, A.B., Haertel, E., Mullis, I.V.S., and Oakes, J. (1989 in press). Assessment in elementary science education. Washington, D.C. National Center for Improving Science Education.

Roberts, S.V. (1989). For the public schools, a mixed report card. New York Times. February 14.

Rogoff, B. and Lave, J. (Eds.) (1984). Every day cognition: Its development in social context. Cambridge, Massachusetts and London, England. Harvard University Press.

Tirozzi, G.N., Baron, J.B., Forgione, P.D., Jr., and Rindone, D.A. (1985). How testing is changing education in Connecticut. Educational measurement: Issues and practice. 4 (2), 12-16.

Whitehead (1929). The aims of education. Cambridge: Cambridge University Press.

DRAFT /Judith Collison

SUPERMARKET PRICE COMPARISONS

SUMMARY: How can you really tell which supermarket will save you the most money? By comparing prices at two supermarkets, students are to design and carry out a study to answer the question.

PRIMARY DEVELOPER: Steven Leinwand (CSDE)

REVIEWERS: Bruce Davey & Judith Collison (CSDE)

COURSE: General Math

GRADE/LEVEL: 9, 10

CURRICULUM TOPIC: Computational skills, using integers, decimals and fractions; understanding units of measurement; conversion between different units; gathering data, organizing data in tabular or graphic form; interpretation of data in problem solving; ability to write conclusion in clear narrative form.

PREREQUISITE KNOWLEDGE:
How and where to shop in the area.

TOOLS: Calculator

SUGGESTED LENGTH OF TIME: 1-2 weeks

IN CLASS: 2-3 class periods

OUTSIDE OF SCHOOL: 2-3 class periods

PARTICIPATION: Mixed (individual and group)

SCORING CRITERIA: Individual and group

ASSIGNMENT

Many local food markets claim to have the lowest prices. But what does this really mean? Does it mean that every item in their store is priced lower, or just some of them? **How can you really tell which supermarket will save you the most money?** Your assignment is to design and carry out a study to answer this question.

Part I (individual)

DESIGN A STUDY

Design a study to determine which supermarket has cheaper prices.Describe your plan in a written report to be handed in. Your report should answer the following questions:

1) What two food markets would you compare?

2) How would you compare the prices at the two stores? How would you select the products to be compared, how would you record the comparisons?

3) How would you analyze the data to be collected, in order to answer the research question?

GROUP MEMBERS: _____
_____
_____
_____

Part II (group)

CONDUCT STUDY

A. The class will be formed into small research groups. You will become a member of a group of 3-4 students. You and the other members of your group will be required to hold two meetings during class time to compare the plans each of you developed, and to develop a final plan for this assignment. Your plan must be designed to answer the three questions listed below, and should specify the responsibilities of each group member:

1) What two food markets would you compare?

2) How would you compare the prices at the two stores? How would you select the products to be compared, how would you record the comparisons?

3) How would you analyze the data to be collected, in order to answer the research question?

B. Your group will hand in your final plan for preliminary grading, comments and suggestions before you carry it out.

C. Your group will then carry out the plan, with each group member doing a separate part of the data collection and analysis. This will probably require two visits to the food markets you have chosen.

D. Each group will separately report on the project. Your final report should include the following information:

1. State the problem the research project was designed to solve
2. State how the data were collected. Include your specific role in collecting the data. (Who did what)
3. State how you analyzed the data. Explain why you chose the method of analysis selected. Include all of the analyses you did.
4. State your final conclusion. In other words, if someone asked you which supermarket has lower prices, how would you answer?
5. Include one or more graphic representations of your data to illustrate and support your final conclusions.

NAME_____

Part III (Individual)

A. Suppose that you are the manager of the store which your group found to be have higher prices. Write a letter to your the group critiquing the study or defending your pricing policies.

B. Advertising is supposed to tell us when and where we can get the best price for things we need and want. It can call our attention of special sales or bargains, but they can also be misleading. Read the article "State Accuses *Fox, Macy's* of False Advertising."

Write an essay on how the stores were trying to mislead the consumers, and what steps consumers can take to avoid being mislead.

NAME_____ S3

Part IV (Individual) TRANSFER TASK

There are many other stores which can be compared using approaches such as you used with the supermarket price comparison. For example, you might want to compare two clothing stores. The task which follows is based on this idea.

Assume that you are a newspaper writer and your editor has asked you and another reporter to do an article comparing two clothing stores - G. Wolf's and J.C. Nickels. A full page advertisement for each of the stores is attached.

You and your reporting partner, Wanda Stark, are going to be writing the article together. She will write the part of the article that deals with quality and service. You are going to be writing the part of the article dealing with price comparisons.

Read and answer the questions below:

1. What information do you need from G. Wolf's and J. C. Nickels in order to do the price comparison? Be very specific - make a list which is complete enough for your assistant, Carruthers, so that he can go to each store and get the information for you.

2. Now assume that Carruthers has let you down. He took your instructions and the pay advance you gave him and never returned. You now have only a half hour to meet your editor's deadline for the article, so you have to use the newspaper ad as your only source of information! Now what do you do?

   Prepare an analysis of the price information in the newspaper article, designed to answer the question - which store has the lowest clothing prices? Attach your computation to your final response. You may use your calculator, but attach and label any numbers you compute which you use in the chart.

   Your answer to question 2 should consist of one or more charts, graphs or tables which you feel help answer the question, "Which clothing store has the lowest prices?" Remember, you are not concerned about quality in your analysis, since Wanda is handling that. Try to design one or more charts, graphs or tables which are very informative as to the question which store has the lower prices, and which are also clear and simple enough for the newspaper reading public to understand. Be sure to label your work so that it is self-explanatory to those who read the newspaper. Then put your name on it and attach it this package.

3. Write a paragraph for the newspaper article which addresses the question of which store has the lowest prices. In writing this paragraph, remember who your audience is and what you trying to communicate to them.

SCORING OF THE TRANSFER TASK: Performance goals and competency indicators

Performance Goal #1	The student can identify appropriate data to gather to solve a practical research problem.
COMPETENCY INDICATORS	To meet the standard for Goal #1, a student's response should show recognition of the fact that an overall answer to the question is not appropriate - breakdowns are needed by male/female clothing and/or types of clothing (suits, jackets. sportswear, etc.)
Performance Goal #2	The student can select and use an appropriate mathematical approach to solve a practical research problem.
COMPETENCY INDICATORS	To meet the standard for Goal #2, a student's response should show that they have applied appropriate methods to analyzing the data in a way that answers the question. An example of an appropriate method is to take the mean of prices for the various categories of clothing (e.g., slacks, pants, jacket, shoes, socks, underwear). To exceed the standard, the student should show some insight into how mathematical sampling methods can focus on a better solution to the problem (example: pricing a full male or female clothing ensemble by computing the average price of pants + shirt + shoes + socks, underwear + coat).
Performance Goal #3	The student can display the data with charts, graphs or tables in a way which helps illuminate the answer to the research question for the reader.
COMPETENCY INDICATORS	To meet the standard for this goal, the student must produce at least one graph, chart or table which is clearly labeled, properly constructed, and addresses the question who has the lower prices. To exceed the standard, the student should produce multiple graphs meeting the basic standard, and/or produce graphs which in the teacher's judgment are especially clever, clear or illuminating of the research question.
Performance Goal #4	The student can draw, from their data, conclusions which are mathematically correct and relevant to the solution of the problem.
COMPETENCY INDICATORS	To meet the standard for Goal #4, the student's response should show recognition of the fact that there are multiple answers to the questions (example: Wolf's prices are lower for men's clothing, but Nickel's are lower for women's clothing). To exceed the standard, the response should contain sub-breakdowns such as "Nickel has lower prices for men's slacks and men's and women's shoes". In either case, the conclusion should follow from and be consistent with the graphs, charts and tables presented.
Performance Goal #5	The student can write his/her conclusion in clear correct narrative form.

NOTES TO THE TEACHER

PRIMARY FOCUS: Mathematical Application

GENERAL OBJECTIVES: CCL Quantitative:

- formulate and solve problems in mathematical terms.
- add, subtract, multiply and divide whole numbers, decimals, fractions and integers.

CCL Mathematics: apply mathematical skills and techniques to complete consumer and job related tasks.

SPECIFIC OBJECTIVES: NCTM Standard :

- apply mathematical modeling to real world situations,
- use ratios
- organize data

PRIOR PREPARATION: Some experience with graphs, tables and comparisons; Discussion of unit costs, average costs, and sales prices

INFORMATION NEEDED: Location of stores, and how to get there.

EQUIPMENT NEEDED: Calculators

SEQUENCING OF IN-CLASS WORK AND HOMEWORK:

TECHNIQUES:

- Suggest that students keep a log of individual work.
- You may wish to introduce the concept of a weighted average: should an item bought on a regular basis, such as bread or milk have the same value in the
- The idea of a typical "market basket" can be introduced.

SUGGESTED TOPICS FOR DISCUSSION:

Fallacies used in misleading advertisements.

TYPES OF REASONING PRIMARILY USED: If this task is administered after students have studied the different types of logical reasoning, the task could be an excellent source of insight, and occasion for assessing their ability to make these distinctions. The examples of each type of reasoning in this task are clear and easily accessible.

1) Deductive: in figuring out unit pricing, then comparing value based on price per unit is deductive

2) Inductive: extending pattern of findings to speculation about the prices of items not considered, is inductive

3) Statistical: using selected items, which are believed to be representative of pricing policy of a particular store, then drawing conclusions about all prices in store based on the sample, is method used in statistical reasoning.

4) Analogical:Using similarities and differences between items, prices and values is reasoning through analogy.

CONNECTIONS:(Related Contexts)

- **to other topics in math**
- **to disciplines outside of math**
- **to students' environment**

SUGGESTED TRANSFER TASKS: (Related Applications)

- **analogous problems: more abstract, or more concrete**
  Shopping for shoes, clothes etc.
  Shopping for tapes
- **interdisciplinary approaches**
  compare cost of seasonal items over time [produce e.g.]
  compare costs at convenience stores vs. supermarkets; With social studies teacher, discuss profit margins, effect of volume, effect of monopoly (if small town, of one store)
- **applications likely to be useful to student**
  Compare area banks for interest on CDs, mortgage rates, or car loan rates. Which is the "best" bank?

G. WOLF'S

Where Quality Rules...

DRESS CLOTHES:

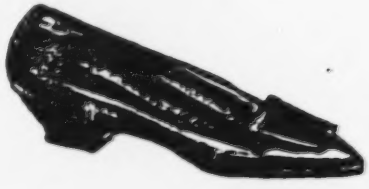
Cameo	Men's Cotton shirts	$ 10 - 18
Cameo	Women's blouses	$ 12 - 20
Breen	Wool skirts	$ 34
Creations	Women's trenchcoats	$ 77
Brown's	Ladies' Cashmere Pantsuit	$ 89
Robertson	Men's Pinstripe Suit	$ 99
Great Western	Men's Leather Jackets	$129
Great Western	Women's Leather Jackets	$149
English Mist	Men's Single breasted topcoat	$169
English Mist	Men's Double breasted topcoat	$189

SPORTS AND CASUAL WEAR:

Chameleon	Pocket t-shirts	$ 8
Hallings	Solid Color sweatshirts (M/W)	$ 10
Speed King	Men's Swimsuits	$ 12 - $ 36
Speed King	Ladies' Swimsuits	$ 14 - $ 39
Hallings	Designer Sweatshirts (M/W)	$ 15
Cameo	Men's Denim shirts	$ 32
Cameo	Ladies' Denim shirts	$ 36
Cody	Denim skirts	$ 38
Heather	Girls' Jeans	$ 40

FOOTWEAR:

Greblinger	Socks	$ 3
McCrawford	Men's dress shoes	$ 49 - $ 89
ProAth	A-1 Model Sneakers (Men's)	$ 59
Gazelle	ACE Model Sneakers (Men's)	$ 79
Biltwell	AAA Model Sneakers (M/W)	$ 79



J. C. NICKELS
Quality Clothes

MEN'S WEAR

Bingo	Pocket t-shirts	$ 11
Cameo	Cotton shirts	$ 10 - 18
Chameleon	Pocket t-shirts	$ 7
English Mist	Men's Single breasted topcoat	$181
English Mist	Men's Double breasted topcoat	$195
Fastfeet	Deluxe	$ 99
Gazelle	ACE Model sneakers	$ 76
Great Western	Men's Leather Jackets	$139
Greblinger	Men's Socks	$ 3
Hallings	Solid Color Sweatshirts	$ 10
Hallings	Designer Sweatshirts	$ 16
Kleeger	Men's Dress Suits	$ 79
ProAth	A-1 Model Sneakers	$ 55
Robertson	Men's Pinstripe Suit	$109
Speed King	Men's Swimsuits	$ 12 - $16

WOMEN'S WEAR

Breen	Wool skirts	$ 34
Brown's	Ladies' Cashmere Pantsuit	$ 89
Cameo	Women's blouses	$ 13 - 21
Cameo	Women's Denim shirts	$ 30
Cody	Denim skirts	$ 38
Creations	Women's trenchcoats	$ 99
Fastfeet	Deluxe Women's sneakers	$ 99
Gazelle	ACE Model sneakers for women	$ 78
Great Western	Women's Leather Jackets	$154
Hallings	Solid Color Sweatshirts	$ 10
Hallings	Designer Sweatshirts	$ 16
Heather	Girls' Jeans	$ 42
Jackson	Ladies' Swimsuits	$ 10 - $25
Speed King	Ladies' Swimsuits	$ 17 - $49
Wanda's	Women's fur-lined boots	$ 39
Wanda's	Women's high heel shoes	$ 16